# The « SentiWeb » Method for Exploring a Database on the Net

Eric Boussard, Antoine Flahault, MD
Le Réseau Sentinelles - INSERM U444 - B3E
Institut fédératif Saint-Antoine de Recherche sur la Santé
Paris, France

*Return of information is one of the main goals of any public health information system. About 25,000 maps and 10,000 graphs may be obtained from the time-series collected in the database of the French Communicable Diseases Network (FCDN). Furthermore, this huge epidemiological atlas is updated each week. What is the optimal way of returning such information? This report discloses the strategies used for enhancing the access facilities to the FCDN database for any users, particularly those without specific training in epidemiology or database query language. The technical options implemented in the SentiWeb server (http://www.b3e.jussieu.fr) are discussed.*

## INTRODUCTION

The French Communicable Diseases Network (FCDN) gathers feed back of information from 1% of the general practitioners (GPs) for the surveillance of several communicable diseases[1]. For 12 years, this system has built a large database with more than 330,000 cases individually described by the sentinel GPs. Going beyond to the surveillance goal, these time series have been the basis for many research programs in epidemiology. For example, we tested the impact of population movements on the time-space dynamic of influenza epidemics[2]. These raw data have motivated the development of methods for epidemiological data presentation[3].

Data are directly entered by GPs on remote terminals and sent to a host computer. After quality control, raw data are statistically processed to provide estimations of incidences. Individual description of case and incidence figures are stored in a relational database (Oracle V7). For each disease under surveillance, incidences are calculated for various period intervals (week, month, quarter, half-year and year), and for various spatial entities (administrative departments, regions, and metropolitan France. To date, more than 800,000 figures of incidences are recorded.

Since September 1995, the Sentinel System has opened a World Wide Web server (http://www.b3e.jussieu.fr/sentiweb). It aims to provide, on the net, a user-friendly tool for exploring this database. It was assumed that the users may have no specific skills in database query language or in epidemiology. Therefore, the input is limited to the use of the mouse and the output is presented as graphs or maps rather than tables or figures.

However, production of maps or graphs obviously requires the use of SQL (Standard Query Language) query, statistic and graphic procedures. Here, two



Figure 1 - The html-form to get time series graph

File    Edit    View    Go    Bookmarks    Options    Directory    Window                    Help

Back    Forward    Home    Reload    Images    Open    Print    Find    Stop

Location: http://www.b3e.jussieu.fr:80/cgi-bin/time_serie?maladie=rougeole&departement=10
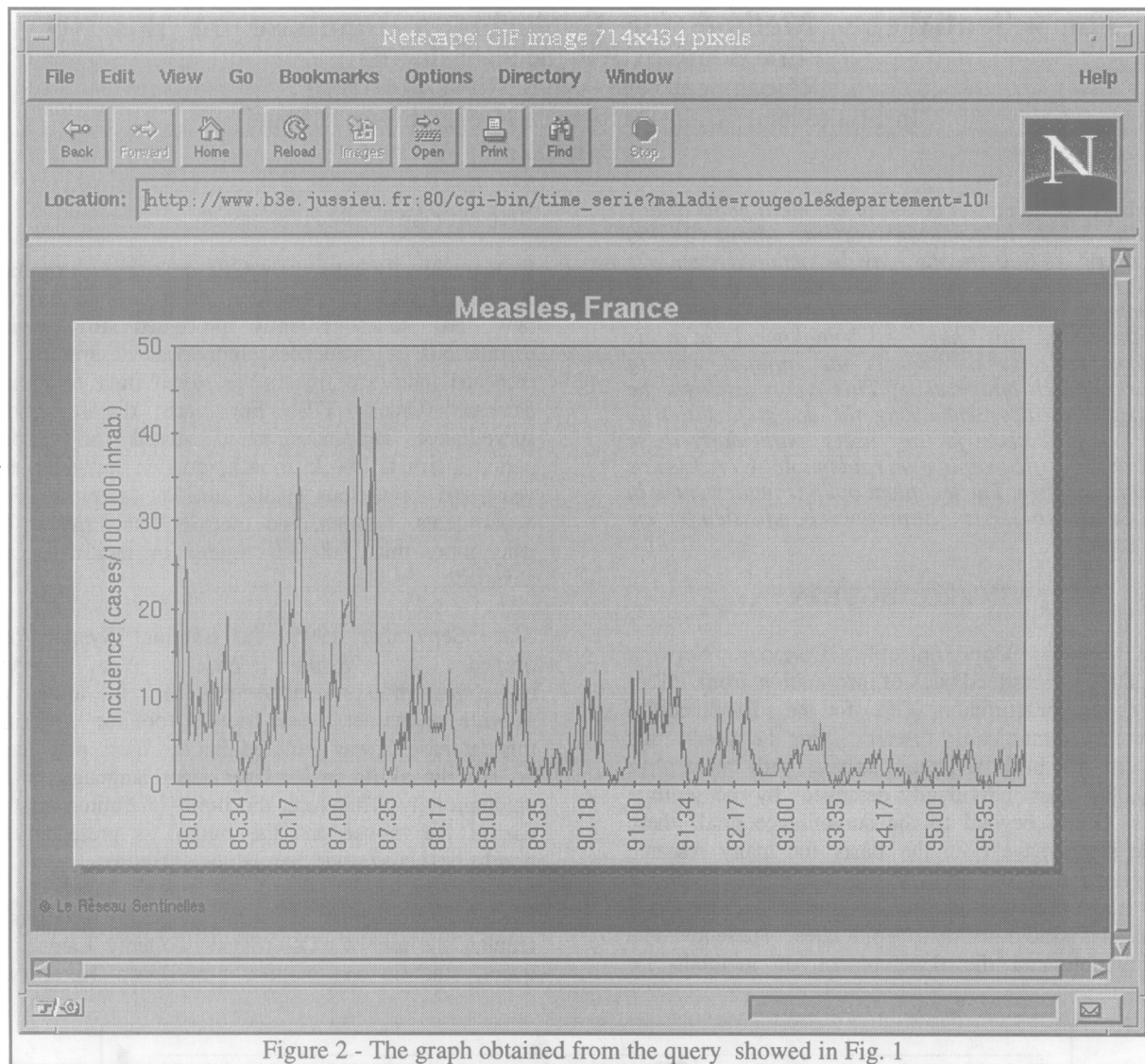


Figure 2 - The graph obtained from the query showed in Fig. 1

examples are described regarding methods for producing on-the-fly graphs and maps.

## AD-HOC PRODUCTION OF GRAPH CHARTS

### Specification of the final output to be obtained

A time series of weekly incidence figures is plotted, for a given disease, in a given geographic area, as a function of time within a specific time range.

### Parameters to be set

The parameters to be set by the user are the disease of interest (influenza, measles, male urethritis, ...), the area code, and the week number starting the time series. In this example, the end date was set to the date of the active user's session.

### Step 1 : HTML form

User preference for the parameters selection is provided on a specific form coded in a HTML (Hyper Text Mark-up Language) page with three scrolled lists (Fig. 1). This page can be interpreted by any web browser. By clicking, the user selects the disease of interest, the administrative area, and the beginning date. In this example, it was the first week of the selected year. All parameters are sent to the server when the user clicks on the button « Send the query ».

### Step 2 : query process

The web server launches a CGI (Common Gateway Interface) program coded in C with the parameter set on the command line. The first module of this program builds the following pre-written SQL query by using the user's parameter set:

670

```
SELECT week, incidence
FROM table_incidence_area_type_by_week
WHERE illness = illness_code
AND area = area_code
AND week > beginning_date
AND week < ending_date
ORDER BY week,
```

This SQL query is sent to the Oracle server which returns data as a time series array of {week, incidence}.

## Step 3 : graphic process

These data are given as input to a graphic module. It uses the graphic resources of commercial graphic software Widget (XrtGraph from KL Group). Widget is a pseudo-object available from the Xwindow interface. The pixmap (raw image) is extracted from the widget display in a virtual screen.

## Step 4 : final output

The pixmap is converted into a GIF format and sent directly to the client by setting the 'Content-Type' to 'image/gif' in the header of the response. The GIF image is directly readable by any web browser. (Fig. 2)

## AD-HOC PRODUCTION OF MAPS

## Specification of the final output to be obtained

Geographical distribution of a disease incidences is mapped from data collected at a local level for a given time period.

## Parameters to be set

The parameters to be set by the user are the disease of interest, the selected time-period, the graphic method to be used.

## Step 1 : HTML form

The HTML specific form to get maps from the database has three scrolled lists and an option menu (Fig. 3). By clicking, the user selects the disease of interest, and the time period. The option menu allows the user to select between two different methods of map drawings : maps with iso-incidences, or maps with filled-in administrative areas. The button « Send the query » sends the user's parameter set to the web server.

## Step 2 : query process

The web server launches the same type of CGI C-code described above. The program begins by building the SQL-query with the user's parameter set to the Oracle server. The SQL query has the following structure:

```
SELECT area, incidence
FROM table_incidence_area_type_time_type
WHERE disease = illness_code
AND date = date_code
ORDER BY area
```

The Oracle server returns information as an array of {area, incidence}.

## Step 3 : statistical process

If the user asked for a map with iso-incidences curves, a statistical process of data is performed. It consists of



Figure 3 - The HTML-form to get a map

671
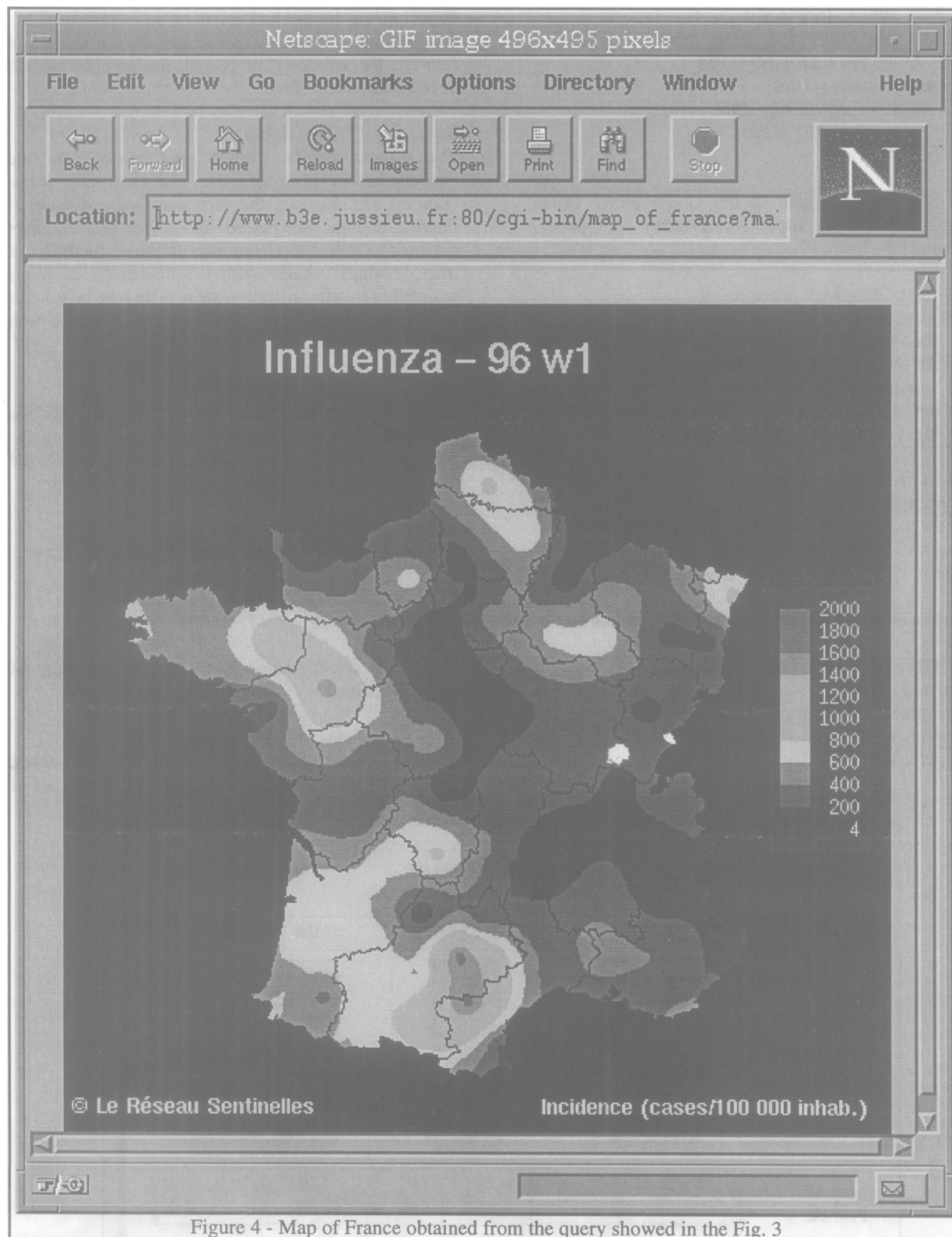
a spatial interpolation by the krigging method.[3] Data distribution, where each incidence value is applied to the geographic coordinates of the main city of the administrative department, is transformed in a regular matrix (here, 100x100), where each value of $(x_{ij}, y_{ij})$ is associated to an estimation of the incidence. The array has then a $\{x_{ij}, y_{ij}, incidence_{ij}\}$ format.

**Step 4 : graphic process**
This array is processed by the graphic module described above. The widget displays the matrix with colorscale coding. The pixmap is extracted and a mask is applied. The mask is a bitmap representing the contour of the metropolitan France, with (or without) the contours of administrative districts, according to the user's preferences.

In case of a simple filling-in areas request, the program fills in the areas in the bitmap mask with a colorscale coding.

**Step 5 : final output**
The resulting pixmap is then converted into a GIF format, sent to, and read directly from, any web browser client (Fig. 4).

## ADVANTAGES AND LIMITS OF THE SENTIWEB METHOD

SentiWeb provides on the net an access to a customizable atlas of more than 25,000 maps and more than 10,000 graphs covering the last 12 years in France. Response times from the user's request to the completion of the display are enough short to allow our methodology, between 15 to 60 seconds for a map and between 5 to 20 seconds for a graph (in our local network).

Producing ad-hoc graphics has many advantages. This strategy avoids the storage of several megabytes of images. Any change in the processing are immediately available for all the atlas. Weekly updates are immediately available on the net. This method obviously consumes computational resources on the server side. Experience shows us that requests are easily satisfied by the server at this time. The web server is a DEC AlphaStation 250 and the Oracle server is a DEC AlphaServer 2100.

The HTML-based user interface has been built to let the user intuitively make a query without the knowledge of the database structure and without the knowledge of a database query language. Since all the query fields could be filled via option menu or scrolled list, a lot of user errors are avoided, such as typing errors or syntax mistakes.

Since September 1995, Sentiweb has served about 9,000 images. Requests came from a lot of different countries in the world, but about 50% from France. The produced graphics are easily understandable for a non-epidemiologist public. The French TV (France 2) downloaded directly some images of the last influenza epidemic from SentiWeb and presented them during the news. In a graph, a peak of incidence may correspond to an epidemic. Maps are color coded for

low or high activity level of incidence of a given communicable disease. Week after week, a user can see the dynamics of the epidemic. Many explanations are provided with graphs and maps obtained on SentiWeb. E-mail may also be sent directly from the web page to the research team in charge of these public health surveillance. Conflicts between the freedom of information access and the risk of spurious interpretation has been discussed in detail elsewhere[4].

## FURTHER DEVELOPMENTS

Remote data entry for regular collection of epidemiological information will be implemented in the future, allowing European (or Global) public health surveillance. Entry tools will also be used soon for assessing the use and the user's satisfaction regarding SentiWeb, in association with regular usage statistics.

Development of client capacities through the use of Java (or 'Java-like') tools will be explored, for enhancing the potential for customization of the user queries. These interfaces could allow to experimention with the remote use of simulation models, such as neural networks, compartmental models, or artificial intelligence programs applied to public health issues.

### References

1. Garnerin P, Saidi Y, Valleron A-J. The French communicable diseases computer network, a seven-year experiment. *Ann NY Acad Sci,* 1992; 670: 29-42.
2. Flahault A, Letrait S, Blin P, Hazout S, Ménarès J, Valleron A-J. Modelling the 1985 influenza epidemic in France. *Stat Med,* 1988; 7: 1147-1202.
3. Carrat F, Valleron A-J. Epidemiologic mapping using the « krigging » method. Application to an influenza-like illness epidemic in France. *Am J Epidemiol,* 1992; 135: 1293-1300.
4. Boussard E, Flahault A, Vibert J-F, Valleron A-J. Sentiweb: Open the French sentinel system database to the world wide public. *British Med J. In press.*